Modeling molecular interactions in proteins and water between quantum chemistery and classical electrostatics.

Rullmann, Johan Antoon Christoffel

Summary

In the last decade tremendous progress has been made in the study of complex heterogeneous systems like solutions and biological macromolecules. Detailed computer modelling of these systems on the atomic level has become feasible, and in fact indispensable, considering the use of simulation techniques in structure determination with X-ray or NMR. Yet the theoretical methods to deal with dynamical systems composed of hundreds of atoms are still in their infancy. In this thesis several aspects of molecular interaction models for biological systems are discussed, with the emphasis on the treatment of electrostatic interactions.

After a short overview of the field is given in the introductory chapter, the theoretical basis of the study of intermolecular interactions is outlined in chapter 2. It is argued that empirical force fields, and many *ab initio* approaches using supermolecules, neglect the serious problem of separating intermolecular interaction energies from the intramolecular energy change due to the presence of other particles. In this way strong interdependencies between potential function parameters are introduced, hampering systematic improvement of these functions and transfer of parameters from *e.g.* crystals to the liquid phase. Conspicuous differences between existing force fields are observed with respect to charge distributions and hydrogen bond strengths. The usual 'effective potential' approach for induction and dielectric effects is invalidated by qualitatively and quantitatively different results obtained with more detailed models. Examples are cited from the literature, showing the sensitivity of calculated molecular energies to variation of the electrostatic model. The conclusion must be that present day force fields are least reliable in situations where the electric field plays a prominent role, such as binding sites, active centres and ionic solutions - cases which are at the heart of many 'molecular engineering' studies. Many of these problems can be attacked by using quantum mechanical perturbation theory to develop force fields. Molecular interactions can be expressed in terms of well defined and calculable monomeric properties. An overview is given of practical approaches to evaluate these quantities, and to represent them in many-body interaction calculations.

In chapter 3 the problem of deriving macromolecular charge distributions from *ab initio* calculations is discussed, with emphasis on the inductive interactions between fragments and on the problem of merging the charge distributions of neighbouring, and partially overlapping, fragments. A model is set up to treat inter- and intra-molecular polarization on equal footing. Helical and extended conformations of poly-glycine, and serine dipeptide are considered as test cases. Especially the long range co-operative polarization of α-helices is represented quite well

by the model.

In chapter 4 *ab initio* charge distributions for protein fragments (amino acid side chains, peptide groups) are derived. The merging of these charge distributions to describe polypeptide chains is investigated in detail, by comparing with *ab initio* results obtained for dipeptides. Polarization plays a small, but distinct role in improving the agreement. The main problem is the correction of fragment distributions for boundary effects. Increasing the overlap between fragments reduces the error. Another problem is to describe the subtle changes when varying the conformation of a group of atoms. This, however, requires a careful analysis of large basis set calculations.

Overall the dipole moments and partial charges from the overlapping-fragments-plus-polarization model are correlated well enough with Hartree Fock SCF results to include mutual polarization of fragments as a standard procedure in macromolecular calculations. An automated 'protein building' program has been developed to retrieve partial charge (or other) data from a data base and to generate the information required by macromolecular energy programs (*cf.* chapter 7).

Chapters 5 and 6 deal with the description of electrostatic interactions in solutions. In chapter 5 a hydration model without periodic boundary conditions is developed, in which the solute and two or three layers of water molecules are embedded in a dielectric continuum. A Monte Carlo program has been written to perform simulations with this model. Subsequently the hydration of alifatic amines in different protonation states has been studied. Experimental energies of protonation can be reproduced quite well, provided that the definition of the dielectric boundary takes into account the van der Waals volume of the enclosed water molecules. The continuum contribution to the energy is found to depend almost exclusively on the the net charge of the solute and on the distance between solute and boundary. Further analysis shows that continuum type models can not represent faithfully displacements of charges and water–water correlations within 3 to 5 Å of the boundary. Improvement of the non-periodic model is only possible by using 'potentials of mean force' for water molecules near the boundary, or by further increasing the number of hydration layers, possibly combined with a simplified molecular description in the outermost layers.

In the amine protonation studies a comparison has also been made between two electrostatic potentials, the semi-empirical GROMOS potential and the *ab initio* TD potential, which is based upon a dipole conserving population analysis. Replacement of one potential by the other has a large effect on the protonation energy in water, and also on protonation energy differences between similar compounds. The hydration energy is sensitive to details of the charge distribution, such as the partial charges of CH_2 and CH_3 groups. Polarizability of water molecules also appears to be essential.

Summary

This is the subject of chapter 6, in which a polarizable water model is developed. After implementing this model in the Monte Carlo hydration program, again protonation energies of amines were calculated. In the case of ionized substituents the agreement with experimental results is much improved by using the polarizable in stead of pair-additive water model. In combination with the polarizable model the *ab initio* potential performs better than the semi-empirical one. A fully self consistent treatment of polarization, allowing for interactions between induced dipoles, is shown to be necessary ánd feasible.

In chapter 7 models from previous chapters are combined in a study of the active site of the cysteine protease papain. Partial charge and polarizability of all protein atoms, as well as a number of crystal waters, were included in the calculations. The central question is whether the active site residues exist as a neutral Cys-His or as a zwitterionic Cys^{-}-His^{+} pair. In vacuum the former is much more stable. However, the surrounding protein and the crystal water near the active site are found to stabilize the zwitterionic state much more than the neutral state. Both states are likely to be in thermal equilibrium. These theoretical results support the interpretation given to experimental data, as well as the proposed mechanism for the catalytic activity of papain, which involves a thiolate anion rather than a thiol group. The main factor causing the stabilization of the ion pair is the field of a long α-helix, but local interactions - involving specific residues and water molecules - turn out to be also important. The dielectric properties of the protein in particular can not be described by a continuum type model. All atom representations are needed to properly balance the screening of charge–charge interactions (which actually may be enhanced in stead of reduced), and the self energy change due to interactions between a charge and its polarizable environment.